Artificial Intelligence for Engineering Design, Analysis and Manufacturing

cambridge.org/aie

Artificial intelligence methods for improving the inventive design process, application in lattice structure case study

Masih Hanifi[1,2], Hicham Chibane[2], Remy Houssin[1], Denis Cavallucci[2] and Naser Ghannad[2]

[1]Strasbourg University, 4 Rue Blaise Pascal, 67081 Strasbourg, France and [2]INSA of Strasbourg, 24 Boulevard de la Victoire, 67000 Strasbourg, France

Research Article

Cite this article: Hanifi M, Chibane H, Houssin R, Cavallucci D, Ghannad N (2022). Artificial intelligence methods for improving the inventive design process, application in lattice structure case study. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* **36**, e20, 1–17. https://doi.org/10.1017/S0890060422000051

Received: 4 July 2021
Revised: 25 January 2022
Accepted: 9 February 2022

Key words:
Artificial intelligence; document embedding; inventive design; lattice structure; similarity computation; text classification

Author for correspondence:
Masih Hanifi,
E-mail: masih.hanifi@insa-strasbourg.fr

Abstract

Nowadays, firms are constantly looking for methodological approaches that help them to decrease the time needed for the innovation process. Among these approaches, it is worth mentioning the TRIZ-based frameworks such as the Inventive Design Methodology (IDM), where the Problem Graph method is used to formulate a problem. However, the application of IDM is time-consuming due to the construction of a complete map to clarify a problem situation. Therefore, the Inverse Problem Graph (IPG) method has been introduced within the IDM framework to enhance its agility. Nevertheless, the manual gathering of essential information, including parameters and concepts, requires effort and time. This paper integrates the neural network doc2vec and machine learning algorithms as Artificial Intelligence methods into a graphical method inspired by the IPG process. This integration can facilitate and accelerate the development of inventive solutions by extracting parameters and concepts in the inventive design process. The method has been applied to develop a new lattice structure solution in the material field.

Introduction

In recent decades, many companies have been competing on reducing the innovation cycle time in order to obtain pioneering profits in the market (Cohen *et al.*, 1996). To achieve this goal, these companies can utilize systematic approaches such as Theory of Inventive Problem Solving (TRIZ). The TRIZ was introduced by Genrich Altshuler. TRIZ can help reduce the time needed to achieve an inventive solution and develop a new product (Altshuller *et al.*, 1996). However, this approach does not provide any means to select among the proposed solutions. Moreover, it does not offer any methods to formulate the problems. To solve these drawbacks, several frameworks have been developed by researchers. Among them, it is worth mentioning the Inventive Design Methodology (IDM). IDM is a TRIZ-based systematic approach that has been developed to complement the TRIZ body of knowledge with other theories such as graph theory (Cavallucci and Strasbourg, 2009). This approach includes the following phases (Zanni-Merk *et al.*, 2011):

(1) Initial Analysis phase: This phase relates to the gathering of the expert's tacit know-how, the knowledge from patents, and other relevant documents. Subsequently, it is time to use tools such as the Problem Graph to transform the collected knowledge into a graphical model (Hanifi *et al.*, 2019).
(2) Contradiction Formulation phase: In this phase, the designers formulate the contradictions, which are physical and technical problems in a system (Chibane *et al.*, 2021).
(3) Solution Concept Synthesis phase: In the third phase, the TRIZ methods such as the contradiction matrix and inventive principles are applied to solve the contradictions.
(4) Solution Concept Selection phase: In the final step, the designers can use an evaluation grid to measure the impact of the concepts (Hanifi *et al.*, 2020*b*).

However, one of the criticisms often leveled is that IDM is time-consuming and lacks the necessary agility (Chai *et al.*, 2005; Cavallucci *et al.*, 2009; Souili and Cavallucci, 2017). This is because a complete map related to a problem situation has to be built at the beginning of the project without considering its effectiveness in solving the problem. As a result, the Inverse Problem Graph (IPG) method was introduced to formulate a problem situation.

The IPG method (Hanifi *et al.*, 2021, 2022) was developed by several authors to formulate a problem situation in the inventive design process. This method consists of seven main steps, which are as follows: (1) Define the project goal, (2) Determine the initial problem, (3) Identify the causes of the initial problem, (4) Rank the identified causes and select the most important one, (5) Determine the type of the selected cause , (6) Extract the formulated contradiction, and (7) Allocate the appropriate parameters to the formulated contradiction. The formulation of contradictions in the IPG method contrasts with the other methods in helping to formulate problems. Indeed, by applying the IPG method, designers search for the causes of a problem at the lower level of a problem situation. In contrast, using methods such as the Problem Graph, designers look for the effect of a general problem at the upper level (Hanifi *et al.*, 2020*a*, 2020*b*). IPG could give the process of IDM the characteristics of agile methodologies, including capacity to generate flexible and rapid response to change and the capability for iterative development (Kumar and Bhatia, 2012). However, the collection of critical data such as parameters and solution concepts within the IDM framework is still done manually, reducing the agility of the process. Hence, it is necessary to integrate automatic information retrieval methods, such as the neural network doc2vec model and machine learning algorithms, into the process to increase its agility.

The doc2vec model was proposed by Le *et al.* (Le and Mikolov, 2014) (paragraph vectors) as a document-embedding method in 2014. This model is applicable in text classification and document similarity calculation (Park *et al.*, 2019). Text classification is known as the task of classifying a given text into a set of predefined classes (Dalal and Zaveri, 2011). This can be done by applying machine learning algorithms (Sarkar, 2019). The main objective of this paper is to integrate the doc2vec model and machine learning algorithms into the inventive design process. Therefore, the main contribution of this paper is to propose a method that facilitates and accelerates the extraction of essential data, helping to develop inventive solutions in the inventive design process.

The rest of the paper is organized as follows. In Section "Literature review", we present the literature on automatic extraction methods, a review of several document-embedding techniques, a method for measuring the similarity, and a description of several machine learning algorithms. Section "Proposed method" displays the structure of the proposed method and describes the steps of this proposal. Then, we evaluate the machine learning algorithms in the section "Evaluation of the reviewed machine learning algorithms". In Section "Application of the proposal to lattice structure case study", a case study is presented in which the proposal is used to formulate the inventive problems related to the lattice structures and extract the solution concepts. In Section "Comparison of the proposal and the problem graph's system", we make a comparison between our proposal and another automatic technique applied in IDM. We present the discussion in the section "Discussion". In the last section, we report the conclusion of the paper.

Literature review

Automatic extraction methods to assist designers

In the literature, various methods have been developed to support the designers in the inventive design process by extracting the information. For example, Han *et al.* (2018b) have proposed a computational tool, based on ontology and analogical reasoning, to assist designers in creative idea generation during the initial stages of inventive design. In addition, the authors in Shi *et al.* (2017) developed an approach using data analytic and text mining techniques to extract design information from engineering perspective. Besides, Chen *et al.* (2019) introduced an approach using data mining and artificial intelligence techniques to bring inspiration in a visual and semantic way. Moreover, in the paper (Song *et al.*, 2019), a data-driven method was developed to build a function co-occurrence network based on the function data in prior product designs. This method helps to detect peripheral and core functions to be included in a product platform. The authors in Siddharth and Chakrabarti (2018) introduced a web-based tool that supports designers in problem-solving by transferring concepts from the biological field to engineering domains. In the paper (Han *et al.*, 2018a), a computational method, based on simulating aspects of human cognition in obtaining combinatorial creativity, was presented. In addition, Sarica *et al.* (2020) have developed a method that uses NLP techniques to extract terms from patents. In the literature, there are also several automatic methods that have been proposed within the IDM framework. Our focus in this paper is on the methods related to the initial phase of IDM.

Automatic extraction tools related to the initial phase of IDM

Generally, it is possible to classify the automatic extraction tools utilized in the first phase of the IDM into two major groups:

Automatic tools to extract information from the patents

Patent documents are an important depository of technical knowledge applied to obtain competitive benefits (Li *et al.*, 2015). To extract patent information, researchers have introduced various patent analysis tools (Valverde *et al.*, 2017). Nevertheless, in this article, we only discuss those that are related to the first phase of the IDM framework. One of the approaches that automatically extracts the IDM concepts was introduced by Souili *et al.* (2015). However, one of the drawbacks of this approach is the extraction of partial solutions and problems from the patents without considering the requirements of designers. Hence, Berduygina and Cavallucci (2020) proposed to use claims hierarchical structure to improve the final output of the IDM-related information extraction tool. This proposal could help removing the repeated information from the extraction. However, drawbacks resulting from the extraction based on the designer's requirements remain in the tool. Another weakness that can be mentioned here concerns the incapability to extract information from scientific papers.

Automatic tools to extract data from scientific papers

Scientific papers include information that their extraction could help improve the quality of human life (Nasar *et al.*, 2018). Therefore, Nédey *et al.* developed a tool within the IDM framework to extract problems, partial solutions, and parameters from scientific articles by improving the IDM patent extraction methodology (Nédey *et al.*, 2018). Nevertheless, this proposal, as its original method, extracts information beyond the designer's requirements, making their analysis quite laborious and time-consuming.

As we have seen, the tools developed under the IDM framework ignore the main requirements of designers to extract the information. This makes analyzing their results time-consuming. For this reason, in this article, we integrate document-embedding techniques and machine learning algorithms into a graphical method inspired by the IPG method. In the next section, we will review two document-embedding techniques, and we will choose one of them to apply to our proposed method.

Document embedding and similarity computation techniques

The Bag of Words (BOW) is a simple method that represents a text as a fixed-length vector (Zhao and Mao, 2018). Nevertheless, this method ignores the meaning of the words, the grammar, and the order in a text (Zhang *et al.*, 2008). Furthermore, BOW does not integrate linguistic meanings either. Therefore, the retrieved information by this method is not understandable (Li *et al.*, 2015). Moreover, extending the number of texts, using the BOW method, leads to high-dimensional and sparse representation. Hence, it is not an effective method to represent the proximity among the texts (Kim *et al.*, 2017).

Word2vec has been proposed by Mikolov *et al.* (2013). This technique permits the calculation of the semantic similarity among two words and to derive similar words semantically (Mimura and Tanaka, 2018). Nevertheless, this method loses this order in a text (Zhang and Zhou, 2019). Hence, to overcome this drawback of word2vec, the researchers developed doc2vec.

Doc2vec is an unsupervised method developed by Le and Mikolov (2014). This method is an extension of word2vec that can express a document as a vector (Aman *et al.*, 2018). Doc2vec provides the possibility to exploit the semantic information existing in a text. Moreover, this method applies to texts of different lengths (Hanifi *et al.*, 2019). This method can also obtain higher accuracy term vectors by extracting the word order information in the text (Zhang and Zhou, 2019). After getting the term vectors of two different texts, the similarity of the terms corresponds to the correlation between their vectors. This similarity can be calculated by a method such as Cosine Similarity (Huang, 2008).

Cosine similarity is a method to measure the similarity among two vectors (two sentences) (Chang *et al.*, 2018). This method helps to find similar sentences to a given text. As a result, we integrated it into our method. To extract the type of each sentence and its related parameters, we also need the machine learning text classification algorithms. In the next section, we will review some of these algorithms.

Text classification and machine learning algorithms

Text classification is defined as the task of classifying a given document into a set of predefined classes according to the extracted features (Dalal and Zaveri, 2011). To automate text classification, there are two main types of machine learning algorithms which include (Sarkar, 2019): (1) Unsupervised learning and (2) Supervised learning. Here, we will focus only on the existing algorithms in the second group, as our proposal uses the provided prelabeled samples to build the model. In what follows, we will describe some of these algorithms.

David Cox developed the logistic regression model in 1958 (Sarkar, 2019). Logistic regression (LR) is a supervised machine learning classification technique based on the probabilistic statistics of the data (Feng *et al.*, 2014). Logistic regression is classified as follows (Park, 2013): (1) Binominal or binary logistic model. (2) Multinomial logistic regression model. Logistic regression is effective in predicting categorical outputs (Kowsari *et al.*, 2019).

The Multilayer Perceptron (MLP) was proposed by Rosenblatt in 1958 (Panchal *et al.*, 2011). The MLP consists of multiple-layers including an input layer, an output layer, and one or more hidden layers between its input and output layers (Ramchoun *et al.*, 2016). This algorithm is applied in the various business and industrial domains to classify and predict problems (Adwan *et al.*, 2014).

Random forest (RF) is an ensemble learning classification algorithm that uses bagging to build multiple decision trees (Singh *et al.*, 2017). Random forest is efficient in handling large datasets and feature sets (Rane and Kumar, 2018). Besides, it does not overfit on large datasets (Pandey *et al.*, 2017). Additionally, it is one of the most accurate algorithms (Silva *et al.*, 2013).

The K-Nearest Neighbor (KNN) was introduced by Cover and Hart in 1968 (Mulak and Talhar, 2015). This algorithm classifies the objects based on nearest training samples in the feature space (Imandoust and Bolandraftar, 2013). The KNN classification algorithm is easy to implement (Pawar and Gawande, 2012). Moreover, it works well even in multi-class document managements (Kowsari *et al.*, 2019). However, KNN requires more time to categorize texts when there is a large number of training samples (Pawar and Gawande, 2012).

The support vector machine (SVM) was introduced by Vapnik (Vapnik, 1995). SVM is a supervised learning algorithm that is based on Structural Risk Minimization (SRM) principle in the statistical learning theory (Lin *et al.*, 2019). SVM is applied to a wide variety of classification problems including nonlinearly separable and high-dimensional problems (Soofi and Awan, 2017). This algorithm generally has the highest classification accuracy (Khan *et al.*, 2010). However, SVM requires more computation time (Khan *et al.*, 2010).

Proposed method

In this section, we integrated doc2vec and machine learning algorithms into a graphical method, inspired by the IPG process, to extract information from existing databases. This enabled us to introduce a new method for the Initial Analysis and Solution Concepts phases of the inventive design process. This method consists of two main parts: (1) Graphical Model to Formulate Contradictions and (2) Agile Automated – Question-Answering System (AA-QAS), which includes the creation of corpora subparts and the question-answering subparts. Figure 1 shows the different parts of this proposal. In the following, we explain these parts in detail.

Part 1: Graphical Model to Formulate Contradictions: We developed a graphical method, inspired by the IPG method, to link the problem formulation process to the second part.

Step 1: Determine the initial problem of IPG: In the first step of the first part, the initial problem is determined by considering the project objective.

Step 2: Find the most relevant causes of the initial problem: Next, it is time to find the most relevant causes (problems) of the initial problem by entering the Agile Automated – Question-Answering System (AA-QAS) in the second step.

Step 3: Select the most important causes and determine their type: At the beginning of the third step, the designers must select the most important causes (problems). Then, it is time to determine their type. (3.a.i) If the selected causes (problems) are harmful-useful, they must be converted into partial solutions. (3.a.ii) In what follows, the causes of the partial solutions should be determined through the application of AA-QAS. (3.a.i′) If the selected causes (problems) are not harmful-useful, the designers should determine their causes by applying AA-QAS.

Step 4: Extract the illustrated contradictions and assign the appropriate parameters: After demonstrating the contradictions on the

https://doi.org/10.1017/S0890060422000051 Published online by Cambridge University Press

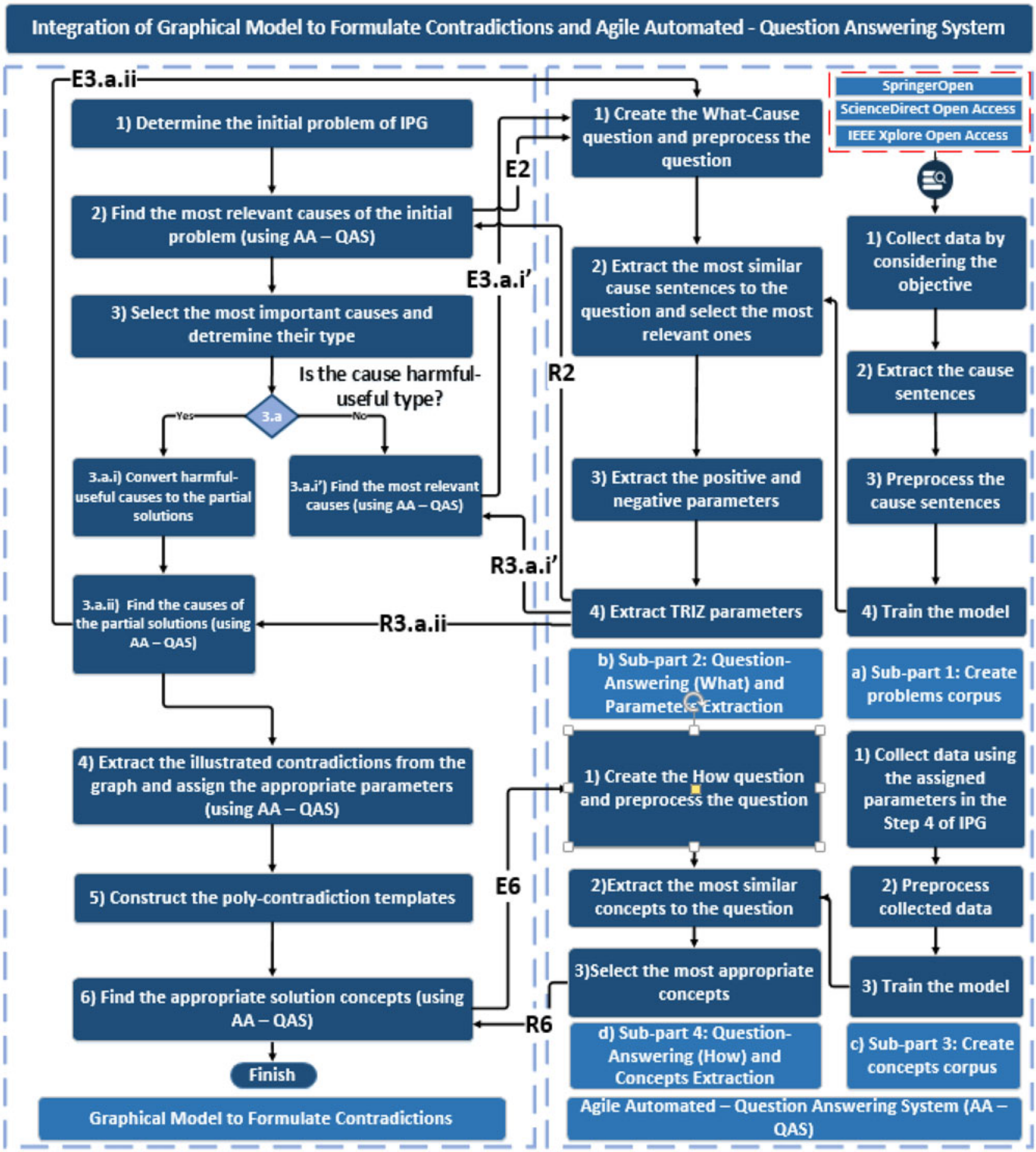


Fig. 1. Integration of modified IPG and agile automated – question-answering system.

graph, it is time to extract them from the graph at the beginning of the fourth step. Then, the parameters are allocated to the contradictions by using the extracted parameters from the corpus.

Step 5: Construct the poly-contradiction templates: In the fifth step, the allocated parameters in the previous step are used to construct the poly-contradiction templates.

Step 6: Find the appropriate solution concepts: At the end of the first part, it is time to extract the solution concepts for the formulated contradictions from the concepts corpus. To do this, it is necessary to enter (E6) subpart 4 to formulate the "How Question". This question helps to extract the concept from the concepts corpus.

Part 2: Agile Automate – Question-Answering System (AA-QAS): This part of the system allows the designers to extract the essential data, including parameters and concepts, from the project corpora. Part 2 consists of the following parts: (a) Creation of the problems corpus, (b) Question-Answering (What) and parameters extraction, (c) Creation of the concepts corpus, and (d) Question-Answering (How) and concepts extraction.

Subpart 2.1: Creation of the problems corpus: In this subpart, the data collection to create the problems corpus and its training are performed. This subpart includes the following steps:

Step 1: Collect data by considering the objective: In the first step of this subpart, it is essential to collect a considerable amount of data considering the area and objective of the project.
Step 2: Extract the cause sentences: Next, the cause sentences are extracted from the corpus created in the first step. To do this, it is necessary to train the doc2vec model and machine learning algorithms by using the provided data samples, including cause and non-cause labels.
Step 3: Preprocess the cause sentences: In the third step, it is necessary to apply certain preprocessing techniques such as Tokenization, Stopword removal, and Lowercase conversion to remove unnecessary content from the data (Sarkar, 2019).
Step 4: Train the model: At the end of this subpart, it is necessary to use doc2vec and the extracted cause sentences to train the model. To implement this training, it is possible to apply the Gensim library (Mimura, 2019).

Subpart 2.2: Question-Answering (What) and Parameters Extraction: The second subpart of the proposal facts as an interface between Part 1 and subpart 2.1. Indeed, this part helps to extract the essential data such as the TRIZ parameters from the problems corpus. The second subpart consists of the following steps:

Step 1: Create a What-Cause question and preprocess the question: At the beginning of the first step of the second subpart, the designers should formulate the question "What causes the problem?" The problem posed by the question could be a partial solution (E3.a.ii), an initial problem (E2), or any other problem (E3.a.i′) that designers require to know its causes. Then, it is essential to use some preprocessing techniques to eliminate undesirable content from the question.
Step 2: Extract the most similar cause sentences to the question and select the most relevant ones: In the second step, the cosine similarity is applied to find the most similar cause sentences to the question. Subsequently, the designers should evaluate them to choose those that are closest to the problem.
Step 3: Extract the positive and negative parameters: After selecting the cause sentences, the system extracts the positive and negative parameters existing in the cause sentences by applying a machine learning algorithm in the third step. For this purpose, the system uses our labeled data to train the machine learning algorithm used. The results of Subpart 2.2 can be used as the causes in Step 2 (R2), Step 3.a.ii (R3.a.ii), or Step 3.a.i′ (R3.a.i′) of Part 1.
Step 4: Extract TRIZ parameters: In the final step of this subpart, TRIZ parameters related to each cause sentence are extracted regarding the positive and negative parameters. To extract these parameters, the system also uses the provided labeled data.

Subpart 2.3: Creation of the concepts corpus: In this subpart, the collection of data to create the concepts corpus and its training are performed. It includes the following steps:

Step 1: Collect data using the assigned parameters in Step 4 of IPG: In the first step, it is essential to collect a considerable amount of data using the assigned parameters in the fourth step of the first part.
Step 2: Preprocess collected data: In the second step, it is necessary to use some preprocessing techniques such as Tokenization, Stopword removal, and Lowercase conversion to remove unnecessary content from the data (Sarkar, 2019).
Step 3: Train the model: At the end of this subpart, it is necessary to use doc2vec to train the model.

Subpart 2.4: Question-Answering (How) and Concepts Extraction: It helps to extract the solution concepts from the concepts corpus. The steps in this subpart are as follows:

Step 1: Create the How question and preprocess the question: The first step begins with the formulation of the "How Questions". To do so, it is possible to use the parameters, including TRIZ, positive, and negative parameters, as the keywords. Next, it is time to apply some preprocessing techniques to remove unwanted content from the formulated questions.
Step 2: Extract the most similar concepts to the question: The cosine similarity could help the designers to find the most similar concepts to the questions in the second step.
Step 3: Select the most appropriate concepts: Finally, the designers must evaluate the concepts in order to select the most appropriate ones. The results (R6) of Subpart 2.4 are used as the solution concepts in Step 6 of Part 1.

Evaluation of the reviewed machine learning algorithms

In this section, we evaluate the accuracy of the reviewed machine learning algorithms for the "Cause, Non-cause" and "Parameters" datasets. Furthermore, we test the ability of the algorithms with the highest precision to predict the labels of several sentences. To perform this evaluation, the system configuration was as follows: Intel Xeon (R) CPU 2.2 GHz, 13 GB RAM in Ubuntu 18.04 environment.

For the first evaluation, we used our "Cause, Non-cause" dataset. Our machine learning algorithm is trained by using this dataset to remove Non-cause sentences from the problem corpus. This dataset consists of 2800 sentences and two labels (61.60% Cause and 38.40% Non-cause). We perform a train-test split using the 80–20 rule where 80% of the data is used for training, and the remaining 20% is applied for testing. The machine learning algorithms reviewed in the literature were trained using the training data and tested on the test set to evaluate their accuracy. For accuracy evaluation, we consider F1, recall, and precision to evaluate the overall accuracy of the machine learning algorithms. Table 1 displays the accuracy of each machine learning algorithm related to the "Cause, Non-cause" dataset.

We apply MLP as the algorithm with the highest precision to predict the labels of five new sentences for the "Cause, Non-cause" dataset. Table 2 shows the result of this application. As the table shows, there is only one error in these predictions. The algorithm predicted non-cause for the sentence N° 4, while its label should be "Cause".

The "Parameters" dataset is used for the second evaluation of machine learning algorithms. This dataset is applied in our process to train our machine learning algorithms, helping to extract the parameters from the sentences. The parameter dataset consists of 3607 sentences and the 4 following groups of parameters: (1)

Table 1. Accuracy of machine learning algorithms related to the "Cause, Non-cause" dataset

Algorithms	F1 scores	Precision scores	Recall scores
Logistic regression	84.37%	84.95%	84.64%
Random forest	82.65%	82.89%	82.85%
K-nearest neighbor	85.28%	85.31%	85.35%
Support vector machine	83.75%	83.96%	83.92%
MLP	85.69%	85.68%	85.71%

Table 2. Results related to the application of MLP algorithm to predict the "Cause, Non-cause" labels of five sentences

N °	Texts	Cause, Non-cause labels
1	Raise the storage pressure can increase the hydrogen density but will also remarkably increase the energy consumption for compression	Cause
2	The compressive strength of CA mortar decreases with higher temperature.	Cause
3	The small increase of the cutting speed value from 30 to 35 m/min leads to an increase of temperature.	Cause
4	Loss of energy in PDE due to the viscosity of the fuel is an important factor.	Non-cause
5	Biodiesel can reduce carbon dioxide (CO_2) emissions.	Cause

The "Positive TRIZ Parameters" group includes 40 labels (39 TRIZ parameters labels + nan label); (2) The "Negative TRIZ Parameters" group consists of 40 labels (39 TRIZ parameters + nan label); (3) The "Positive Parameters" group has 78 labels; and (4) The "Negative Parameter" group includes 51 labels. To evaluate the reviewed machine learning algorithms by the parameters dataset, we first selected eight highest labels in each group of parameters as follows:

(1) "Positive TRIZ Parameters" group (469 sentences, Parameters: 15.77% Strength, 14.07% Object-generated harmful factors, 13.21% Temperature, 12.36% Use of energy by stationary object, 11.72% Loss of energy, 11.30% Speed, 10.87% Quantity of substance, 10.66% nan (nan means that there is not positive TRIZ parameter in the sentence), …).
(2) "Negative TRIZ Parameter" group (410 sentences, Parameters: 13.90% Strength, 13.41% Loss of energy, 12.92% Temperature, 12.68% Loss of substance, 12.43% Reliability, 11.95% Use of energy by stationary object, 11.46% Object-generated harmful factors, 11.21% nan (nan means there is not negative TRIZ parameter in the sentence),…).
(3) "Positive Parameters" group (327 sentences, Parameters: 13.76% Weight, 13.45% Energy consumption, 13.14% Energy efficiency, 12.84% Thermal conductivity, 12.53% CO2 emission, 11.92% Density, 11.62% Thermal stability, 10.70%, nan (nan means that there is not a positive parameter in the sentence), …).
(4) "Negative Parameters" group (230 sentences, Parameters: 7.80% Energy consumption, 7.80% Temperature, 7.80% Loss of energy, 7.56% Compressive strength, 7.56% Weight, 5.85% Energy dissipation, 5.85% Material loss, 5.85% nan (nan means that there is not a negative parameter in the sentence),…).

Table 3. Evaluation of machine learning algorithms for the "Positive TRIZ Parameter" group

Algorithms	F1 scores	Precision scores	Recall scores
Logistic regression	83.35%	85.24%	83.67%
Random forest	78.57%	81.42%	79.59%
K-nearest neighbor	91.81%	92.12%	91.83%
Support vector machine	80.50%	85.84%	81.63%
MLP	78.62%	81.79%	79.59%

Table 4. Evaluation of machine learning algorithms for the "Negative TRIZ parameter" group

Algorithms	F1 scores	Precision scores	Recall scores
Logistic regression	80.60%	84.09%	80%
Random forest	87.83%	89.77%	88%
K-nearest neighbor	84.66%	88%	84%
Support vector machine	84.27%	86.66%	84%
MLP	80.66%	83.42%	80%

Table 5. Evaluation of machine learning algorithms for the "Positive Parameter" group

Algorithms	F1 scores	Precision scores	Recall scores
Logistic regression	86.36%	87.09%	87.10%
Random forest	81.38%	83.22%	80.64%
K-nearest neighbor	76.64%	81.52%	77.41%
Support vector machine	83.93%	86.45%	83.87%
MLP	87.52%	88.38%	87.09%

Then, we performed a train-test split using the 80–20 rule where 80% of data is used for training, and the remaining 20% is used for testing. In what followed, the reviewed machine learning algorithms in the literature were trained using the training data and tested on the test to verify their accuracy. Tables 3, 4, 5, and 6 show the accuracy for each machine learning algorithm related to the parameter dataset.

We applied the algorithms with the highest precision to predict the labels related to four groups of the parameter dataset for four sentences. Tables 7 and 8 illustrate the result of these predictions. As shown in the tables, there is one error for the

Table 6. Evaluation of machine learning algorithms for the "Negative Parameter" group

Algorithms	F1 scores	Precision scores	Recall scores
Logistic regression	82.96%	87.77%	83.33%
Random forest	74.14%	78.88%	75%
K-nearest neighbor	74.24%	85.71%	75%
Support vector machine	83.33%	88.88%	83.33%
MLP	91.75%	93.33%	91.66%

"Positive Parameters" group in Table 8. The machine learning algorithm we used predicted "nan" for sentence N°4, while its label should be "CO_2 emission".

Application of the proposal to the lattice structure case study

In this section, we applied the proposed method "Integration of Graphical Model to Formulate Contradictions and Agile Automated – Question-Answering System" to the Lattice Structure (LS) case study. This application helps us to evaluate its applicability. Due to the wide use of LS in energy-absorbing applications, the energy absorption of this kind of structure has always been an interesting research topic for materials scientists and engineers (Li *et al.*, 2019; Edouard *et al.*, 2021). Energy-absorbing structures are the components that convert kinetic energy into other types of energy, such as plastic strain through large deformations of the material (Fazilati and Alisadeghi, 2016). Figure 2 illustrates a lattice structure. In this case study, we will identify the factors that affect the energy absorption of the lattice structure and extract the possible solution concepts. To do so, we first created the problems corpus by collecting about 7000 articles. We, then, applied the proposed method to formulate the inventive problems related to the lattice structure case and extract their possible concepts. In the following, we will explain the steps of this application.

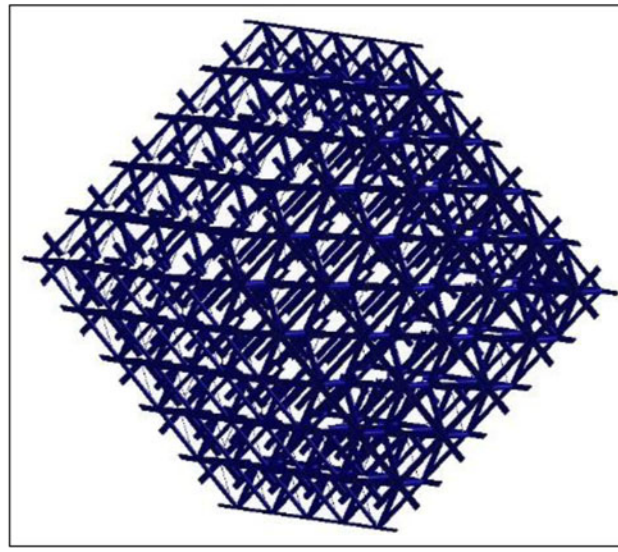


Fig. 2. An example of a lattice structure.

First part

In the first part of the case study, we created a problem database by following these steps: At the beginning of the first step of this

Table 7. Results related to the application of the machine learning algorithms with the highest accuracy to predict the "Positive and Negative TRIZ Parameters" labels of four sentences

N°	Texts	Positive TRIZ Parameters	Negative TRIZ Parameters
1	Raise the storage pressure can increase the hydrogen density but will also remarkably increase the energy consumption for compression.	Quantity of substance	Use of energy by stationary object
2	The compressive strength of CA mortar decreases with higher temperature.	nan	Strength
3	The small increase of the cutting speed value from 30 to 35 m/min leads to an increase of temperature.	nan	Temperature
4	Biodiesel can reduce carbon dioxide (CO_2) emissions.	Object-generated harmful factors	nan

Table 8. Results related to the application of the machine learning algorithms with the highest accuracy to predict the "Positive and Negative Parameters" labels of four sentences

N°	Texts	Positive Parameters	Negative Parameters
1	Raise the storage pressure can increase the hydrogen density but will also remarkably increase the energy consumption for compression.	Density	Energy consumption
2	The compressive strength of CA mortar decreases with higher temperature.	nan	Compressive strength
3	The small increase of the cutting speed value from 30 to 35 m/min leads to an increase of temperature.	nan	Temperature
4	Biodiesel can reduce carbon dioxide (CO_2) emissions.	nan	nan

https://doi.org/10.1017/S0890060422000051 Published online by Cambridge University Press

Table 9. Part of the similar cause sentences to the first formulated question

	Cause Sentences	Similarity	Question
1	It is also evident that there is a monotonic enhance in the energy absorption with the density of the EPS foam and the relative density of the octet-truss LS.	0.871	What reduces the energy absorption of the lattice structure?
…	…	…	…
710	This is because the strong rotation will make the structure easy to have local vibration, but at the same time, the strong chirality is more likely to yield a small porosity, result in the decrease of photonic bandgap.	0.701	What reduces the energy absorption of the lattice structure?

part, we chose “Improvement of the energy absorption of the Lattice Structure” as the objective. Then, we downloaded about 7000 articles, including 231,360 sentences, using “Lattice Structure” as the keyword, from the ScienceDirect Open Access, SpringerOpen, and IEEE Xplore Open Access data sources. After collecting the data, we extracted the cause sentences in the second step. To do so, we first trained our doc2vec model by applying the prepared data sample, including “Cause and Non-cause” labels. Then, we used MLP as one of the evaluated machine learning algorithms to extract the cause sentences. In the third step, we used the standard Natural Language Toolkit (NLTK) python library to perform the preprocessing of the extracted cause sentences. We trained the doc2vec model by applying the extracted cause sentences in the last step of this part. This helps to use cosine similarity to extract the similar cause sentences to the question from the corpus.

Second part

In the second part of the case study, we formulated the contradictions related to the lattice structure by following the steps defined in Subpart 2.2 of the AA-QAS and Part 1:

In the first step of this part, we determined the initial problem, which was that “The energy absorption of the lattice structure is reduced”, by considering the project objective.

After defining the initial problem, we had to determine the causes of this problem in the second step. Therefore, we entered the AA-QAS to use the information in the database. For this purpose, we first formulated our question, which was “What reduces the energy absorption of the lattice structure?” Then, we preprocessed the question by lemmatization, lowercase conversion, and tokenization. In what followed, we used the most similar method to find the cause sentences nearest to our formulated question. Table 9 displays some of these sentences. We should also mention that the minimum threshold for this case study is considered to be 0.70.

Subsequently, we analyzed the proposed cause sentences and selected the most relevant ones, as shown in Table 10, for the defined initial problem of the Inverse Problem Graph.

We extracted the parameters, including the positive, negative, and TRIZ parameters in the third step of the second part. To do so, we first trained our machine learning algorithm. Subsequently, this algorithm allowed us to extract the parameters. Table 10 also shows the extracted parameters related to the selected causes.

In the fourth step, we first wrote them down in an appropriate format and linked them to the initial problem, as Figure 3 illustrates. In this figure, the first cause “The volume fraction of the lattice structure is enhanced” was inspired by the first row of Table 10, which is “when the volume fraction is reduced from 40.4% to 12.5%, the maximum energy absorption efficiency enhances from 35.20% to 45.95%”. In the same way, we formulated the other causes of the initial problem based on the other rows of Table 10. Then, we should select the most important causes. Here, we selected just one cause, which is “The volume fraction of the lattice structure has been enhanced”, in order to show the application of our method. Next, we needed to determine the type of selected cause by asking the question “Is the cause harmful-useful type?”. In this case study, our selected cause “The volume fraction of the lattice structure has been increased” was a harmful-useful problem. Hence, we had to convert this problem to the partial solution “To enhance the volume fraction of the lattice structure”, which means that we converted the problem structure to the partial solution structure, as Figure 3 demonstrates.

Table 10. The selected cause sentences related to the first formulated question and their extracted parameters

N°	Selected Cause Sentences	Positive Parameters	Negative Parameters	TRIZ Parameters
1	When the volume fraction is reduced from 40.4% to 12.5%, the maximum energy absorption efficiency enhances from 35.20% to 46.95%.	Energy absorption	nan	Reliability, Shape
2	The 316 L stainless steel Gyroid lattice shows higher energy absorption capacity and stiffness compared to the BCC lattice.	Energy absorption, Stiffness	nan	Reliability, Shape, Strength
3	Compare to BCC structure and diamond, tetrakaidecahedron structure is able to absorb high energy at high stress.	Energy absorption	nan	Reliability, Shape
4	Most hollow truss lattice have high energy absorption and strength capability than solid truss lattice due to the higher specific second moment of inertia of the constituent beam.	Energy absorption, Strength	nan	Strength, Reliability, Shape
5	The energy absorption of the Gyroid LCS is increased by increasing the relative density.	Energy absorption	nan	Reliability, Shape

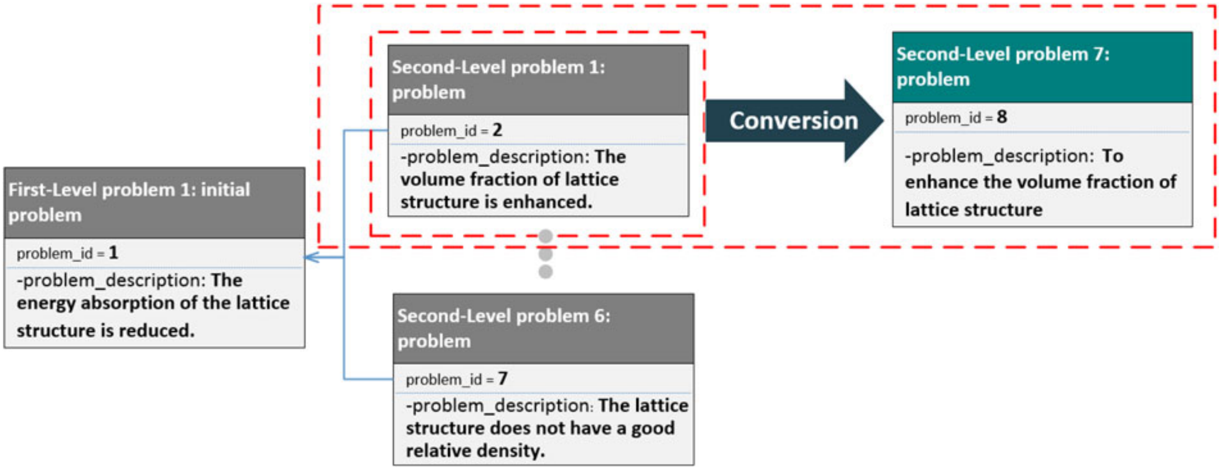


Fig. 3. Connection of the causes to the initial problem, selection of the most important causes, and conversion of the selected causes to the partial solution.

Table 11. Part of the similar cause sentences to the second question

	Cause Sentences	Similarity	Question
1	The strength of the scaffold enhances with the diameter of the dense core, as the porosity is decreased.	0.882	What enhances the volume fraction of the lattice structure
…	…	…	…
1255	Porous carbon materials are attracting the attention of researchers largely due to their excellent electronic conductivity, confined nanospace, high specific area, and high structural stability.	0.703	What enhances the volume fraction of the lattice structure

Once the selected cause was converted to the partial solution, we formulated the following question "What enhances the volume fraction of the lattice structure" at the beginning of the fifth step. We then preprocessed the question, and we extracted the cause sentences most similar to the question. Table 11 shows some of the extracted cause sentences from the database.

In the sixth step, we selected the most probable causes of partial solutions and extracted their parameters by applying our machine learning algorithm, as shown in Table 12. As the first row in the table shows, the system also proposed the sentences that include "porosity", being one of the nearest words to volume fraction.

We interpreted the selected causes of the partial solution to write them on the graph in the seventh step. Figure 4 illustrates the graphical model of this case study.

In the eighth step, we extracted the illustrated contradiction for the selected cause from the graphical model, as shown in Figure 6. The contradiction in this image is between the energy absorption of the structure and its strength. This means that attempting to enhance the strength results in reducing energy absorption. Figure 5 illustrates two different structures, one with a high energy absorption, Figure 5b, and the other with a high strength, Figure 5a. As shown in Figure 5a, the structure with a high strength and low energy absorption leads to the rebound of the object. Conversely, as shown in Figure 5b, the structure with a high energy absorption and low strength is deformed to absorb the kinetic energy.

In the ninth step, we assigned the evaluation parameters to the initial problem and the cause of the partial solution, and the action parameter to the partial solution, as Figure 6 demonstrates. To do so, we used the extracted parameters by our system. These parameters help us to extract the concepts from our database. In this allocation, we allocated two parameters to the energy absorption, as Figure 6 displays. The first one is "Reliability of the structure" and the second one is "Shape of the structure".

Table 12. The selected causes related to the second question and their extracted parameters

N°	Selected Cause Sentences	Positive Parameters	Negative Parameters	TRIZ Parameters
1	The strength of the scaffold enhances significantly with the diameter of the dense core, as the porosity is decreased.	Strength	nan	Strength, Strength
2	While porous T64 scaffold can decrease stiffness due to their enhanced porosity.	nan	Stiffness	Strength
3	For a small volume fraction, the impact of a rough surface is relatively higher than for a high volume fraction, resulting in lower strength.	nan	Strength	Strength
4	It is evident that stiffness enhances with enhance in volume fraction.	Stiffness	nan	Strength

https://doi.org/10.1017/S0890060422000051 Published online by Cambridge University Press

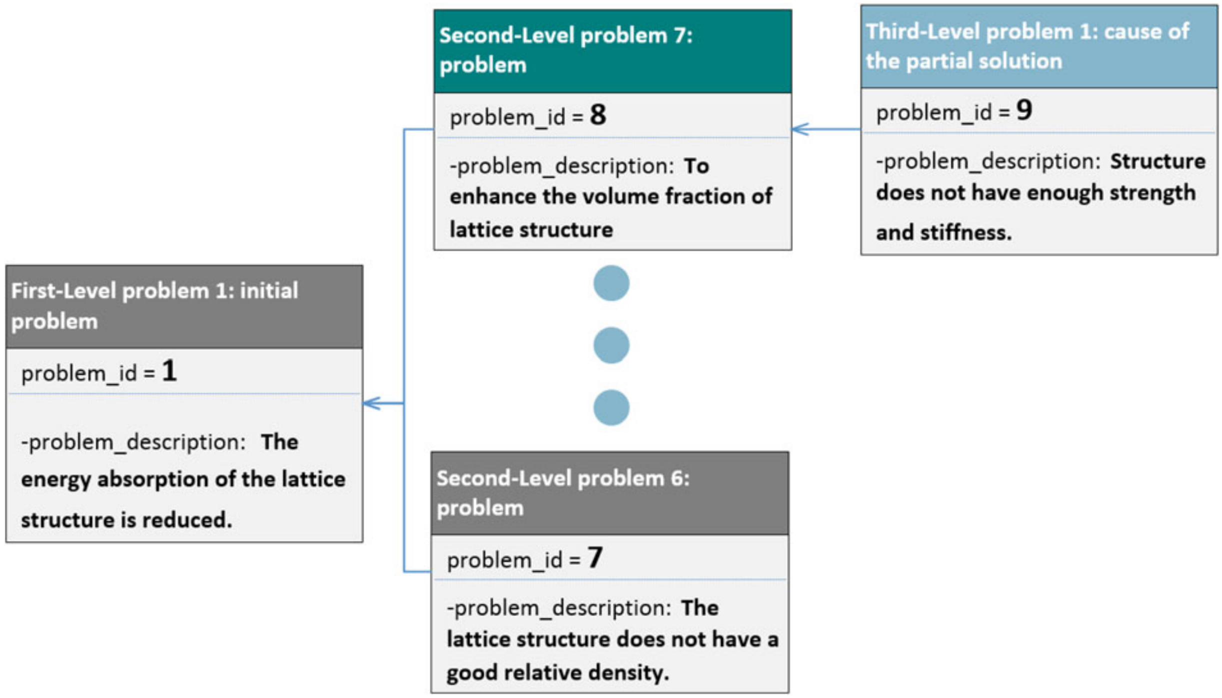


Fig. 4. Graphical model of the case study.

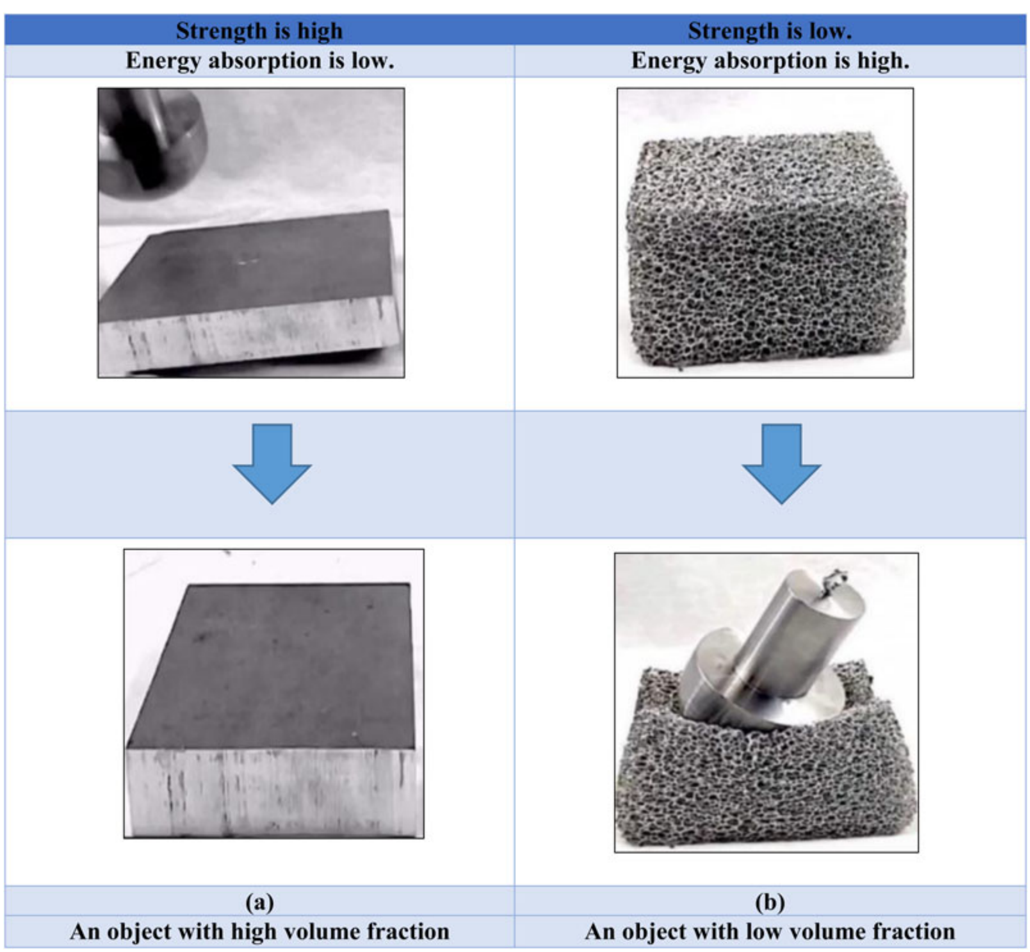


Fig. 5. Comparison of the strength and energy absorption of two structures with a low and high volume fraction ("Duocel® Foam Energy Absorbers").

https://doi.org/10.1017/S0890060422000051 Published online by Cambridge University Press

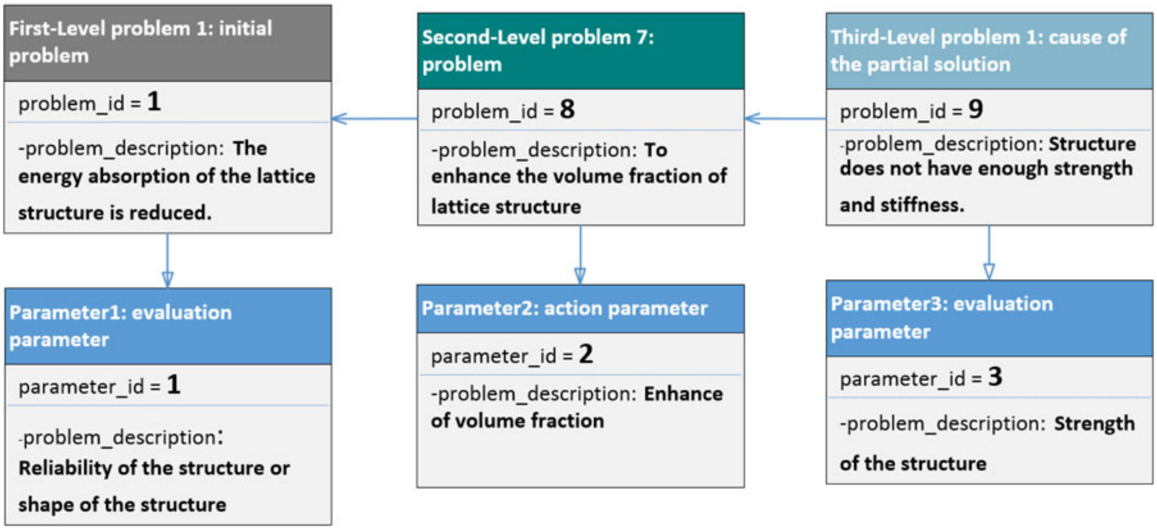


Fig. 6. Allocation of parameters to the extracted contradiction.

In the final step of the second part, we should apply the assigned parameters in the previous step of the proposal to construct our poly-contradiction table, as presented in Figure 7.

Third part

In the third part of our case study, we created a concepts database by using the assigned parameters of the ninth step of the case study. This part consists of the following steps:

In the first step of the third part, we downloaded the articles from the scientific data sources by using the assigned parameters, including energy absorption and strength, in the tenth step of the second part as the keywords. We used the standard Natural Language Toolkit (NLTK) python library to implement the preprocessing of the extracted data in the second step. In the last step of the third part, we trained the model by applying doc2vec.

Fourth part

In the fourth part of the case study, we extracted the solution concepts from the created concepts database through the following steps:

In the first step of this part, we formulated the question "How is it possible to increase the energy absorption (positive parameter) without reducing the strength (negative parameter) of the lattice structure?". This question helped us to extract the solution concepts from the concepts database in the second step of the fourth part. In the final step, we analyzed the extracted concepts to select the most appropriate ones. Table 13 displays the extracted concepts by the system. These concepts help us to develop our solutions.

At the end of this case study, we combined the first (application of the arch geometry) and the fourth concept (application of glass fiber and carbon fiber as the material) to develop our solution. We were inspired by the first concept "The ARCH lattice structures have superior energy absorption and mechanical properties" to add arch geometry to the interior surface of our lattice structure. Besides, the fourth concept helped us to use glass fiber as the material of one of the layers of the structure. This concept also served us to apply Onyx (a mix of carbon fiber and plastic) to construct other layers. Figure 8 displays the different views of our proposed lattice structure.

Comparison of the proposal and the problem graph's system

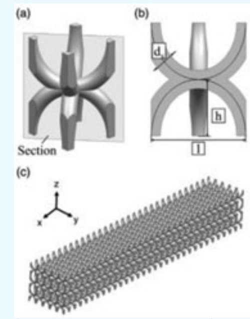
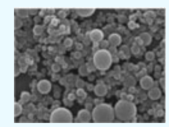
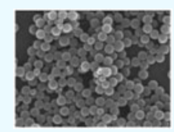
In this section, we compare our proposal with the conventional system based on the Problem Graph method.

To analyze the IPG's automated system, we selected five articles related to the extracted sentences for the Lattice Structure

	Enhance of volume fraction	
	Low	High
Strength of the structure	☹	☺
Reliability or shape of the structure (energy absorption is enhanced – deformation is enhanced)	☺	☹
Shape of the structure: The appearance of a system.		
Strength of the structure: The extent to which the object is able to resist changes in response to force.		
Reliability of the structure: The ability of the structure to perform its intended function in a predictable manner.		

Fig. 7. Construction of poly-contradiction by applying the parameters.

Table 13. The extracted solution concepts from the concepts database for the formulated contradiction

N°	Solution concepts	Similarity	Images
1	The ARCH lattice structures have superior energy absorption and mechanical properties (Ding *et al.*, 2021).	0.8880	 (Ding *et al.*, 2020, 2021)
2	The composition ratio of micro and nano silica particles contributed to the significant effects on the dynamic stiffness, stress transmissibility, and energy absorption performance (Sekiguchi, 2017).	0.8501	 Micro silica particles (Devi *et al.*, 2016)  Nano silica particles (Devi *et al.*, 2016)
3	They believed that the energy absorption capacity of rubber concrete could be accurately reflected using the normalized energy absorption value of compressive strength, which was 54%–79% higher than that of ordinary concrete (Hu *et al.*, 2021).	0.7850	 Rubber (Bala *et al.*, 2014)
4	Moreover, dispersed chopped glass and carbon fibers were used to produce fiber-reinforced polymer concrete with enhanced strength, stiffness, and energy absorption.	0.7730	 Glass fiber (Muley *et al.*, 2015)  Carbon fiber (Muley *et al.*, 2015)
5	The auxetic lattice reinforced composites with a unique combination of stiffness and energy absorption (Li *et al.*, 2018).	0.7605	 Auxetic lattice (Li *et al.*, 2018)

case study. As shown in Table 14, the system extracted one cause sentence from each of the selected articles by considering the need of the designer (formulated question). Considering that the analysis of each sentence requires 5 min, the analysis of each of the articles also took 5 min ($1 \times 5 = 5$ min). Therefore, the average time to analyze one article through the application of IPG's automated system is 5 min, as illustrated in the table.

Table 15 illustrates the information about the capability of the Problem Graph's automated system in extracting the elements from the article. To construct this table, we used the same article

https://doi.org/10.1017/S0890060422000051 Published online by Cambridge University Press

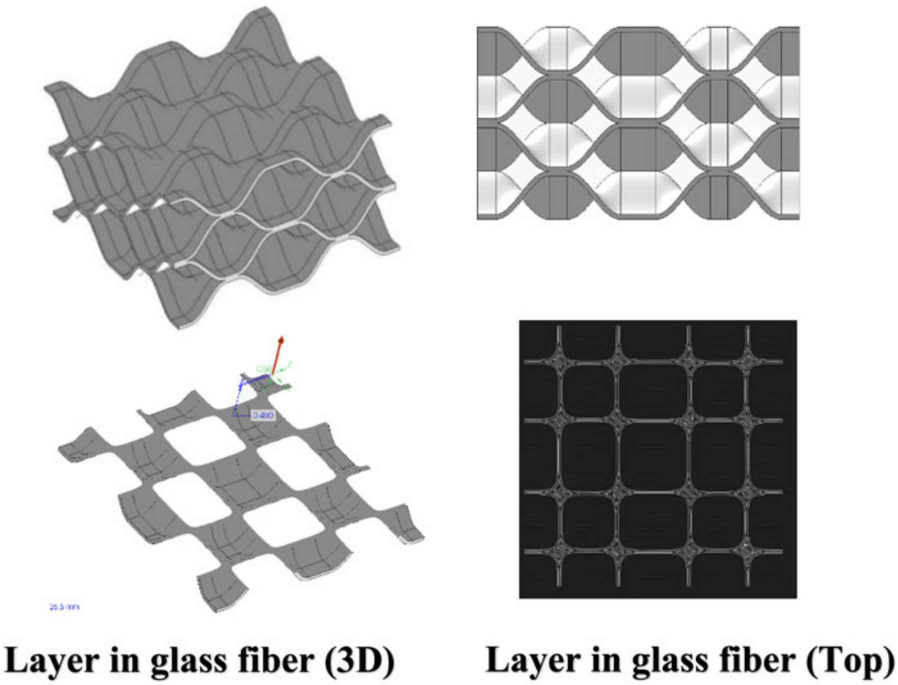


Fig. 8. The proposed solution for the lattice structure.

Table 14. Analysis of the IPG's automated system

Inverse Problem Graph's Automated System			
Article	Number of extracted cause sentence (Considering the need)	Time required to analyze sentence	Total time required to analyze one article
Article 1	1	5 min	5 min
Article 2	1	5 min	5 min
Article 3	1	5 min	5 min
Article 4	1	5 min	5 min
Article 5	1	5 min	5 min
Average	**1**	**5 min**	**5 min**
Keyword used to extract	**Lattice structure**		

Table 15. Analysis of the conventional system based on the problem graph

Problem Graph's system			
Article	Number of extracted elements (Without considering the need)	Time required to analyze each element	Total time required to analyze one article
Article 1	5	5 min	25 min
Article 2	17	5 min	85 min
Article 3	7	5 min	35 min
Article 4	30	5 min	150 min
Article 5	9	5 min	45 min
Average	**13.6**	**5 min**	**68 min**
Keyword used to extract	**Lattice structure**		

https://doi.org/10.1017/S0890060422000051 Published online by Cambridge University Press

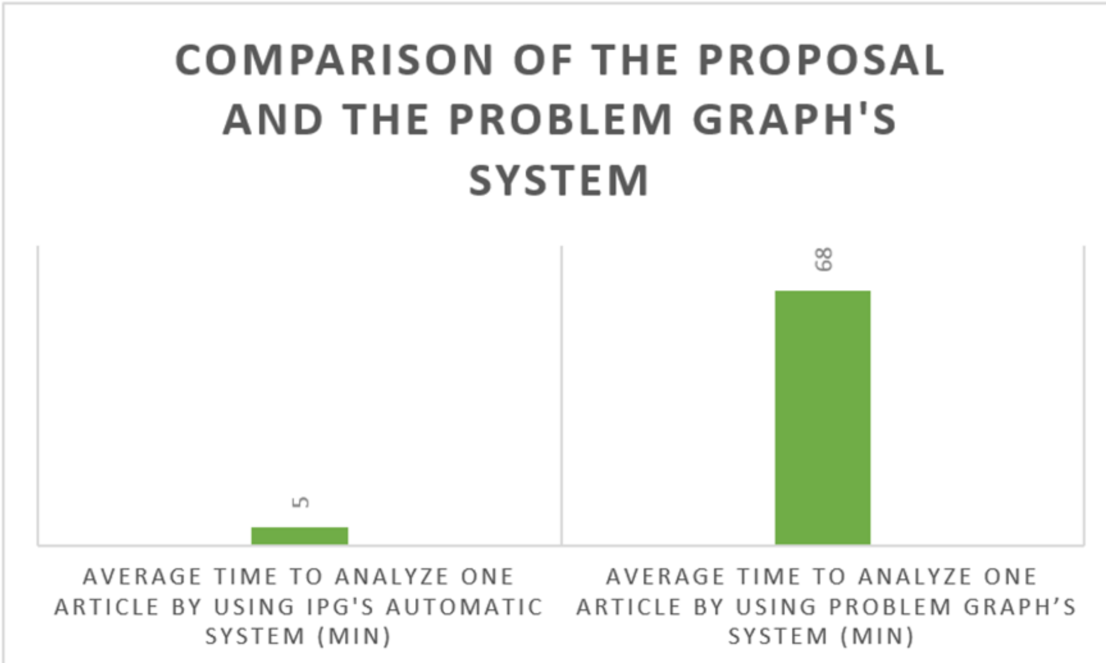


Fig. 9. Comparison of the proposal and the problem graph's system in average time to analyze one article.

as the one used to analyze the IPG's automated system. As shown in Table 15, the number of retrieved elements for the first article, without considering the designer's requirement, was five. Considering that the analysis of each element required 5 min, the analysis of the first article took 25 min. The automatic system retrieved 17 partial solutions and problems from the second article. Hence, the analysis of the whole article required 85 min. Table 15 shows the number of extractions for the other three articles, as well as the time required for their analysis. The table also displays the average time (68 min) required to analyze the extracted data from the conventional system.

Here, we compared the collected information in Tables 14 and 15. As shown in the "Number of extracted cause sentences" column in Table 14, the IPG's automated system extracted the cause sentences from the articles by using its question-answering part. This part allows the system to take into account the needs of designers in each extraction, thus reducing the total time required to analyze one item. In contrast, as the column "Number of extracted elements" in Table 15 illustrates, the conventional system extracts all the elements from the article without considering the needs of the designers. This increases the analysis time of one article in this system. Figure 9 shows a comparison between the average time to analyze one article in the IPG's automated system and the conventional system.

Discussion

In this paper, we presented an automated method to retrieve the information such as parameters and solution concepts. The contribution of this work to the inventive design process is reflected in several aspects. First, the use of similarity computation in the automated system can help to extract similar cause sentences to the formulated questions from scientific data. This capability of the system can facilitate and accelerate the collection of data by the designers in the initial analysis phase of inventive design process. As the second contribution, the system can extract the parameters, including positive, negative and TRIZ parameters, from the cause sentences by using machine learning text classification algorithms. This can significantly reduce the amount of time and effort required to formulate a problem situation in the process. As the third, the use of assigned parameters to the formulated problems enables the designers to extract the nearest concepts to the formulated problems. The existence of this capability in the process can facilitate and accelerate the development of inventive solutions for the designers. As the last contribution, we can mention that the application of our proposal helped to design a new lattice structure in the case study section of this article. This new structure can improve the energy absorption of materials, which can be used for shock absorption and cushioning applications.

Our analysis of these initial results shows some limitations that we would also like to underline. The first is that the formulation of the questions has an impact on the information retrieved by the system. Hence, if the system does not receive an appropriate question, it will not be able to propose acceptable sentences, used as the problems or the concepts. One of the solutions to this drawback could be the integration of an automatic system that proposes questions to the designers. Secondly, this system does not provide any technique to evaluate and select the extracted concepts. To solve this drawback, it is essential to integrate the method into the process that makes the evaluation and the selection of the proposed concepts possible. The third drawback is that the system uses supervised machine learning algorithms, which learn from data samples and their associated training labels. Providing a complete data sample to help the system extracts the parameters is time-consuming. In this case, using transfer learning in future works can reduce the amount of labeled data and resources needed to train new models.

Conclusion

In this study, we developed a new method for extracting essential information, including parameters and solution concepts, in the inventive design process. To propose our method, we first reviewed some of the leading document-embedding techniques to highlight their advantages and limitations. As such, we selected doc2vec, which can extract semantic and order information from text. Next, we analyzed one of the most common techniques, called cosine similarity, to measure the similarity among extracted vectors by doc2vec. In what followed, we reviewed several machine learning algorithms. In the end, we proposed to integrate doc2vec, cosine similarity, and the machine learning algorithms into a graphical method to develop a new approach. We then tested the capability of the approach to collect essential parameters and concepts by applying it to the case study of Lattice Structure. Based on this application, we realized that our proposal could facilitate and accelerate the formulation of contradictions and the extraction of concepts in the inventive design process.

Further research is necessary to appreciate our proposal. As the first one, in order to facilitate question formulation, it is possible to integrate a question/suggestion system, proposing the relevant questions to the designers through the application of the keywords related to a formulated contradiction. Secondly, another future study might be to develop a method to facilitate the evaluation of concepts for the designers.

References

Adwan O, Faris H, Jaradat K, Harfoushi O and Ghatasheh N (2014) Predicting customer churn in telecom industry using multilayer preceptron neural networks. *Life Science Journal* **11**, 75–81.

Altshuller G, Al'tov G and Altov H (1996) *And Suddenly the Inventor Appeared: TRIZ, the Theory of Inventive Problem Solving*. Worcester, MA: Technical Innovation Center, Inc.

Aman H, Amasaki S, Yokogawa T and Kawahara M (2018) A Doc2vec-based assessment of comments and its application to change-prone method analysis. In *2018 25th Asia-Pacific Software Engineering Conference (APSEC)*, Nara, Japan, pp. 643–647. doi:10.1109/APSEC.2018.00082.

Bala A, Sehgal VK and Saini B (2014) Effect of fly ash and waste rubber on properties of concrete composite. *Concrete Research Letters* **5**, 842–857.

Berduygina D and Cavallucci D (2020) Improvement of automatic extraction of inventive information with patent claims structure recognition. In Arai K Kapoor S and Bhatia R (eds), *Intelligent Computing*, Vol. **1229**. Cham: Springer International Publishing, pp. 625–637. doi:10.1007/978-3-030-52246-9_46.

Cavallucci D and Strasbourg I (2009) From TRIZ to inventive design method (IDM): towards a formalization of inventive practices in R&D departments. *Innovation* **18**, 2.

Cavallucci D, Rousselot F and Zanni C (2009) Assisting R&D activities definition through problem mapping. *CIRP Journal of Manufacturing Science and Technology* **1**, 131–136. doi:10.1016/j.cirpj.2008.09.014.

Chai K-H, Zhang J and Tan K-C (2005) A TRIZ-based method for new service design. *Journal of Service Research* **8**, 48–66. doi:10.1177/1094670505276683.

Chang W, Xu Z, Zhou S and Cao W (2018) Research on detection methods based on Doc2vec abnormal comments. *Future Generation Computer Systems* **86**, 656–662. doi:10.1016/j.future.2018.04.059.

Chen L, Wang P, Dong H, Shi F, Han J, Guo Y, Childs PR, Xiao J, Wu C (2019) An artificial intelligence based data-driven approach for design ideation. *Journal of Visual Communication and Image Representation* **61**, 10–22. doi:10.1016/j.jvcir.2019.02.009.

Chibane H, Dubois S and De Guio R (2021) Innovation beyond optimization: application to cutting tool design. *Computers & Industrial Engineering* **154**, 107139. doi:10.1016/j.cie.2021.107139.

Cohen MA, Eliasberg J and Ho T-H (1996) New product development: the performance and time-to-market tradeoff. *Management Science* **42**, 173–186. doi:10.1287/mnsc.42.2.173.

Dalal MK and Zaveri MA (2011) Automatic text classification: a technical review. *International Journal of Computer Applications* **28**, 37–40. doi:10.5120/3358-4633.

Devi G, Al Balushi DALR, Ahmed SJ and Almawali NS (2016) Synthesis and application of nano and micro-silica particles to enhance the mechanical properties of cement concrete. *Concrete Research Letters* **7**, 113–122.

Ding R, Yao J, Du B, Zhao L and Guo Y (2020) Mechanical properties and energy absorption capability of ARCH lattice structures manufactured by selective laser melting. *Advanced Engineering Materials* **22**, 1901534. doi:10.1002/adem.201901534.

Ding R, Yao J, Du B, Li K, Pan M, Zhao L and Guo Y (2021) Flexural properties of ARCH lattice structures manufactured by selective laser melting. *Advanced Engineering Materials* **5**, 2001440. doi:10.1002/adem.202001440.

"Duocel® Foam Energy Absorbers". Available at http://ergaerospace.com/applications/duocel-foam-energy-absorbers/.

Edouard R, Chibane H and Cavallucci D (2021) New characterizing method of a 3D parametric lattice structure. *FME Transactions* **49**, 894–895. doi:10.5937/fme2104894E.

Fazilati J and Alisadeghi M (2016) Multiobjective crashworthiness optimization of multi-layer honeycomb energy absorber panels under axial impact. *Thin-Walled Structures* **107**, 197–206. doi:10.1016/j.tws.2016.06.008.

Feng J, Xu H, Mannor S and Yan S (2014) Robust logistic regression and classification. *Advances in Neural Information Processing Systems* **27**, 253–261.

Han J, Shi F, Chen L and Childs PRN (2018a) The combinator – a computer-based tool for creative idea generation based on a simulation approach. *Design Science* **4**, e11. doi:10.1017/dsj.2018.7.

Han J, Shi F, Chen L and Childs PRN (2018b) A computational tool for creative idea generation based on analogical reasoning and ontology. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* **32**, 462–477.

Hanifi M, Chibane H, Houssin R and Cavallucci D (2019) Improving inventive design methodology's agility. In *New Opportunities for Innovation Breakthroughs for Developing Countries and Emerging Economies*, Vol. **572**. Cham: Springer International Publishing, pp. 216–227. doi:10.1007/978-3-030-32497-1_18.

Hanifi M, Chibane H, Houssin R and Cavallucci D (2020*a*) A method to formulate problem in initial analysis of inventive design. In Nyffenegger F, Ríos J, Rivest L and Bouras A (eds), *Product Lifecycle Management Enabling Smart X*, Vol. **594**. Cham: Springer International Publishing, pp. 311–323. doi:10.1007/978-3-030-62807-9_25.

Hanifi M, Chibane H, Houssin R and Cavallucci D (2020*b*) Contribution to TRIZ in combining lean and inventive design method. In Cavallucci D, Brad S and Livotov P (eds), *Systematic Complex Problem Solving in the Age of Digitalization and Open Innovation*, Vol. **597**. Cham: Springer International Publishing, pp. 280–291. doi:10.1007/978-3-030-61295-5_23.

Hanifi M, Chibane H, Houssin R and Cavallucci D (2021) IPG as a new method to improve the agility of the initial analysis of the inventive design. *FME Transactions* **49**, 549–562. doi:10.5937/fme2103549H.

Hanifi M, Chibane H, Houssin R and Cavallucci D (2022) Problem formulation in inventive design using Doc2vec and cosine similarity as artificial intelligence methods and scientific papers. *Engineering Applications of Artificial Intelligence* **109**, 104661. doi:10.1016/j.engappai.2022.104661.

Hu S, Tang H and Han S (2021) Energy absorption characteristics of PVC coarse aggregate concrete under impact load. *International Journal of Concrete Structures and Materials* **15**, 1–16. doi:10.1186/s40069-021-00465-w.

Huang A (2008) Similarity measures for text document clustering. *Proc. Sixth N. Z. Comput. Sci. Res. Stud. Conf. NZCSRSC2008*, Vol. **4**. Christch. N. Z., pp. 9–56.

Imandoust SB and Bolandraftar M (2013) Application of K-nearest neighbor (KNN) approach for predicting economic events: theoretical background. *International Journal of Engineering Research and Applications* **3**, 605–610.

Khan A, Baharudin B, Lee LH and Khan K (2010) A review of machine learning algorithms for text-documents classification. *Journal of Advances in Information Technology* **1**, 4–20. doi:10.4304/jait.1.1.1-1.

Kim HK, Kim H and Cho S (2017) Bag-of-concepts: comprehending document representation through clustering words in distributed representation. *Neurocomputing* **266**, 336–352. doi:10.1016/j.neucom.2017.05.046.

Kowsari K, Jafari Meimandi K, Heidarysafa M, Mendu S, Barnes L and Brown D (2019) Text classification algorithms: a survey. *Information* **10**, 150. doi:10.3390/info10040150.

Kumar G and Bhatia PK (2012) Impact of agile methodology on software development process. *International Journal of Computer and Electronics Engineering* **2**, 46–50.
Le Q and Mikolov T (2014) Distributed representations of sentences and documents. *International Conference on Machine Learning* **32**, 1188–1196.
Li M, Ming X, He L, Zheng M and Xu Z (2015) A TRIZ-based trimming method for patent design around. *Computer-Aided Design* **62**, 20–30. doi:10.1016/j.cad.2014.10.005.
Li T, Chen Y, Hu X, Li Y and Wang L (2018) Exploiting negative Poisson's ratio to design 3D-printed composites with enhanced mechanical properties. *Materials & Design* **142**, 247–258. doi:10.1016/j.matdes.2018.01.034.
Li D, Liao W, Dai N and Xie YM (2019) Comparison of mechanical properties and energy absorption of sheet-based and strut-based gyroid cellular structures with graded densities. *Materials* **12**, 2183. doi:10.3390/ma12132183.
Lin Y, Zhu X, Zheng Z, Dou Z and Zhou R (2019) The individual identification method of wireless device based on dimensionality reduction and machine learning. *The Journal of Supercomputing* **75**, 3010–3027. doi:10.1007/s11227-017-2216-2.
Mikolov T, Sutskever I, Chen K, Corrado GS and Dean J (2013) Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems* **26**, 3111–3119.
Mimura M (2019) An attempt to read network traffic with Doc2vec. *Journal of Information Processing* **27**, 711–719. doi:10.2197/ipsjjip.27.711.
Mimura M and Tanaka H (2018) Leaving all proxy server logs to paragraph vector. *Journal of Information Processing* **26**, 804–812. doi:10.2197/ipsjjip.26.804.
Mulak P and Talhar N (2015) Analysis of distance measures using K-nearest neighbor algorithm on KDD dataset. *International Journal of Science and Research* **4**, 2101–2104.
Muley P, Varpe S and Ralwani R (2015) Chopped carbon fibers innovative material for enhancement of concrete performances. *International Journal of Applied Science and Engineering* **1**, 164–169.
Nasar Z, Jaffry SW and Malik MK (2018) Information extraction from scientific articles: a survey. *Scientometrics* **117**, 1931–1990. doi:10.1007/s11192-018-2921-5.
Nédey O, Souili A and Cavallucci D (2018) Automatic extraction of IDM-related information in scientific articles. *International TRIZ Future Conference* **541**, 213–224. doi:10.1007/978-3-030-02456-7.
Panchal G, Ganatra A, Kosta YP and Panchal D (2011) Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *International Journal of Computer Theory and Engineering* **3**, 332–337. doi:10.7763/IJCTE.2011.V3.328.
Pandey N, Sanyal DK, Hudait A and Sen A (2017) Automated classification of software issue reports using machine learning techniques: an empirical study. *Innovations in Systems and Software Engineering* **13**, 279–297. doi:10.1007/s11334-017-0294-1.
Park H-A (2013) An Introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing* **43**, 154–164. doi:10.4040/jkan.2013.43.2.154.
Park EL, Cho S and Kang P (2019) Supervised paragraph vector: distributed representations of words, documents and class labels. *IEEE Access* **7**, 29051–29064. doi:10.1109/ACCESS.2019.2901933.
Pawar PY and Gawande SH (2012) A comparative study on different types of approaches to text categorization. *International Journal of Machine Learning and Computing* **2**, 423–426. doi:10.7763/IJMLC.2012.V2.158.
Ramchoun H, Amine M, Idrissi J, Ghanou Y and Ettaouil M (2016) Multilayer perceptron: architecture optimization and training. *The International Journal of Interactive Multimedia and Artificial Intelligence* **4**, 26–30. doi:10.9781/ijimai.2016.415.
Rane A and Kumar A (2018) Sentiment classification system of Twitter data for US Airline Service Analysis. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. **1**. Tokyo, Japan, pp. 769–773. doi:10.1109/COMPSAC.2018.00114.
Sarica S, Luo J and Wood KL (2020) Technet: technology semantic network based on patent data. *Expert Systems with Applications* **142**, 112995. doi:10.1016/j.eswa.2019.112995.
Sarkar D (2019) *Text Analytics with Python: a Practitioner's Guide to Natural Language Processing*. Berkeley, CA: Apress. doi:10.1007/978-1-4842-4354-1.
Sekiguchi Y (2017) Effects of mixed micro and nano silica particles on the dynamic compressive performances of epoxy adhesive. *Applied Adhesion Science* **5**, 1–12. doi:10.1186/s40563-017-0083-y.
Shi F, Chen L, Han J and Childs P (2017) A data-driven text mining and semantic network analysis for design information retrieval. *Journal of Mechanical Design* **139**, 111402. doi:10.1115/1.4037649.
Siddharth L and Chakrabarti A (2018) Evaluating the impact of idea-inspire 4.0 on analogical transfer of concepts. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* **32**, 431–448. doi:10.1017/S0890060418000136.
Silva LOLA, Koga ML, Cugnasca CE and Costa AHR (2013) Comparative assessment of feature selection and classification techniques for visual inspection of pot plant seedlings. *Computers and Electronics in Agriculture* **97**, 47–55. doi:10.1016/j.compag.2013.07.001.
Singh A, Halgamuge MN and Lakshmiganthan R (2017) Impact of different data types on classifier performance of random forest, naïve Bayes, and K-nearest neighbors algorithms. *International Journal of Advanced Computer Science and Applications* **8**, doi:10.14569/IJACSA.2017.08 1201.
Song B, Luo J and Wood K (2019) Data-driven platform design: patent data and function network analysis. *Journal of Mechanical Design* **141**, 021101. doi:10.1115/1.4042083.
Soofi A and Awan A (2017) Classification techniques in machine learning: applications and issues. *Journal of Basic & Applied Sciences* **13**, 459–465. doi:10.6000/1927-5129.2017.13.76.
Souili A and Cavallucci D (2017) Automated extraction of knowledge useful to populate inventive design ontology from patents. *TRIZ – The Theory of Inventive Problem Solving*. Cham: Springer International Publishing, pp. 43–62. doi:10.1007/978-3-319-56593-4_2.
Souili A, Cavallucci D, Rousselot F and Zanni C (2015) Starting from patents to find inputs to the problem graph model of IDM-TRIZ. *Procedia Engineering* **131**, 150–161. doi:10.1016/j.proeng.2015.12.365.
Valverde UY, Nadeau J-P and Scaravetti D (2017) A new method for extracting knowledge from patents to inspire designers during the problem-solving phase. *Journal of Engineering Design* **28**, 369–407. doi:10.1080/09544828.2017.1316361.
Vapnik VN (1995) *The Nature of Statistical Learning Theory*. New York, NY: Springer New York. doi:10.1007/978-1-4757-2440-0.
Zanni-Merk C, Cavallucci D and Rousselot F (2011) Use of formal ontologies as a foundation for inventive design studies. *Computers in Industry* **62**, 323–336. doi:10.1016/j.compind.2010.09.007.
Zhang H and Zhou L (2019) Similarity judgment of civil aviation regulations based on Doc2Vec deep learning algorithm. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Suzhou, China, pp. 1–8. doi:10.1109/CISP-BMEI48845.2019.8965709.
Zhang W, Yoshida T and Tang X (2008) Text classification based on multi-word with support vector machine. *Knowledge-Based Systems* **21**, 879–886. doi:10.1016/j.knosys.2008.03.044.
Zhao R and Mao K (2018) Fuzzy bag-of-words model for document representation. *IEEE Transactions on Fuzzy Systems* **26**, 794–804. doi:10.1109/TFUZZ.2017.2690222.

Masih Hanifi is a Post-Doctoral Researcher at INSA of Strasbourg, Strasbourg, France. His research interests are in the fields of Industrial Engineering, Inventive Design, and Artificial Intelligence.

Hicham Chibane is an Assistant Professor at INSA of Strasbourg, Strasbourg, France. His research interests include Design Engineering and Mechanical Engineering.

Remy Houssin is a Professor at the University of Strasbourg, Strasbourg, France. His research interests include Industrial Engineering and Mechanical Engineering.

Denis Cavallucci is a Professor at INSA of Strasbourg, Strasbourg, France. His research interests include Industrial Design, Industrial Engineering, Mechanical Engineering, and Artificial Intelligence.

Naser Ghannad is a Ph.D. candidate at INSA of Strasbourg, Strasbourg, France. His research interests include Artificial Intelligence and Industrial Engineering.